Center for Independent Experts (CIE) Independent Peer Review of the Marine Recreational Information Program (MRIP) Fishing Effort Survey (FES) Calibration Model

Cynthia M. Jones
Director, Center for Quantitative Fisheries Ecology
Old Dominion University

For The Center of Independent Experts (CIE)

August 2017

Table of Contents

Executive Summary

The task of the MRIP Calibration Review Panel was to evaluate the performance of a new calibration model developed by F. Jay Breidt, Teng Liu, and Jean D. Opsomer of Colorado State University that permits conversion of telephone-survey effort to mail-survey effort and vice versa. The review of the MRIP FES Calibration took place at the Sheraton Silver Springs, in Silver Springs, MD on June 27-29, 2017. Dr. Paul Rago chaired the meeting which included three reviewers from the CIE (Ali Arab, Robert Hicks, Cynthia Jones) and three representing the Fisheries Management Councils and ASMFC (Jason McNamee, Fredric Serchuk, Patrick J. Sullivan).

A survey of recreational fishing effort has been conducted through a random-digit dial (RDD) telephone survey of coastal county households (CHTS) since 1981. With the advent of caller ID, portable prefixes and the proliferation of wireless-only households, the response rate has fallen below 10%. NMFS has chosen a mail survey (FES) to replace the CHTS after a three-year period from 2015-2017 with both surveys overlapping. The calibration model has been applied to the first year (2015) and five, two-month waves of the second year (2016) that has been completed of that overlapping period.

The proposed calibration model is based on a modification of the Fay-Herriot small area estimation method. The Fay-Herriot method (Fay and Herriot, 1979) is well established in the statistical literature and has known statistical behavior. Drs. Breidt and Opsomer and Mr. Liu modified the variance estimation component of that method to be analytically tractable and readily programmed in widely available software. It is fit as a log-normal model regressed on population size and state-by-wave factors with data from the 17 states along the US Atlantic and Gulf coasts. The differences in the non-sampling errors (e.g. frame coverage differences) were modeled with available covariates such as wireless coverage. The difference in the estimates includes the effect of sampling with different survey methods and an “irrational” factor that includes trends over time that could not be explicitly identified as influential covariates. Although some of the differences in effort estimation could be attributed to the increase in wireless only households, the majority of the difference could not be explained with existing available data. As the next year and one half of data become available, the MRIP team will have an opportunity to cross validate the model and evaluate the stability of model parameters. The Panel report includes recommendations to do so. After much consideration, the Panel concurred that this was an appropriate model for calibration.

Although the Fay-Herriot small-area estimation method is well suited for the CHTS to FES calibration, other approaches exist. The statistical team has examined modifications to their approach. For example, through use of the Akaike Information Criteria (AIC), they were able to determine that a simple time-varying ratio estimate that included error performed poorly compared with the current model. The modelers tested Bayesian approaches, but none were presented at the meeting.

TOR1e requested that the panel comment on the accuracy of the CHTS and the FES, but this is not possible for several reasons. The main reason is that anglers self-report their trip number in

surveys that occur off the fishing grounds and there is no external validation of effort by an unbiased observer. Anglers must recall the number of trips that they took within the past two months when asked in the mail or telephone surveys. Many anglers do not keep a diary, although perhaps some keep a calendar, but there is a possibility that these trips are mis-remembered. While there may be little motivation to exaggerate fishing effort, a variety of factors can result in the reported trips differing from the actual number of trips taken and this type of problem is well documented in the survey literature. To measure accuracy one must undertake special surveys that match off site reports with on-site observations and this is best done in small area surveys. Because the effort estimate is combined with CPUE from the on-site angler intercept survey (APAIS) to estimate catch, there is an advantage to the fact that the FES is more efficient, statistically sound, and can potentially have a larger sample size. Larger sample size (more respondents) often results in smaller variance and better characterization of the effort distribution and, thus may result in less uncertainty when combined to produce estimates of catch.

In TOR2, we were asked to comment on the proceedings and issues around them, thus addressing process. I concur with the panel report (Appendix 4).

Having just completed the NAS MRIP Review, and having participated heavily in reviewing the FES and APAIS methodologies, had read much of the literature surrounding the survey methodologies, I was very familiar with the issues underlying the review of the calibration model. However, I noticed that several important reviews, reports, and manuals hadn't been posted for the panel. I and fellow panelists requested these materials on the first day of the meeting and they were promptly made available on the Confluence website. Moreover, the statisticians were not aware of the TORs until shortly prior to the meeting and had less time to prepare their presentations to address the TORs directly. Although they were able to provide us with additional information and presentations by the second day, it would have been better aligned if they had more notice.

During the meeting, I brought up my concerns with communication to the angling public about the calibration model and why the survey method was being changed. I have found that conveying ideas such as a random sample to the lay public challenging even for a trained communicator. These ideas are not simple and the FES is complex. A recent article in the Virginian Pilot by our local outdoor writer complained that NMFS was transitioning to an old-fashioned survey method and why didn't they just use smartphones (Tolliver, 2017)? The difficulty of the task of communicating to the angling public shouldn't be underestimated.

Communication to stock assessment scientists and fishery managers is also vital as the transition to the new survey is completed. The marked difference in effort estimates between the FES and CHTS has ramification of assessment of stock status, how to knit the time-series together, and on the allocation of catch between the commercial and recreational sectors. In some fisheries, the initial impact will be large and possibly disruptive. As time passes and the new survey estimate time series grows longer, problems may diminish. In the meantime, MRIP communication to these two

groups will also rely on the difficult task of conveying concepts that underlie survey sampling, an area of statistics not commonly taught even to quantitative scientists.

Background

To develop a survey of recreational fishing, the location of the fishing area and the length of the season must be considered. For the coastal US, marine recreational fishing is extensive in area, covers both public and private access, and can occur year round on a variety of species and gears. One of the appropriate survey types for such a challenging assessment is a *complemented* survey, wherein effort is assessed off site of the fishery and catch-per-unit effort (CPUE) is observed directly on site. Both the Marine Recreational Fishery Statistics Survey (MRFSS) and the MRIP are two types of complemented surveys. MRFSS uses a telephone survey (Coastal Household Telephone Survey, CHTS) to measure effort off site and the Access-Point Angler Intercept Survey (APAIS) to obtain CPUE on site. In contrast, MRIP uses a mail survey, the Fishing Effort Survey (FES) to obtain effort offsite and APAIS for CPUE onsite. The changeover from the CHTS to the FES has resulted in significant differences in estimates of effort that must be reconciled as a new time series of effort is established. The review that I was asked to participate in was to evaluate a model to calibrate effort between the CHTS and FES. Dr. Opsomer noted in his presentation that when other large surveys in the US had change their survey methods, that they didn't try to establish a calibration between the old and new survey methods, so the NMFS MRIP calibration is one of the first of its kind.

Since 1981 the NMFS has monitored recreational fishing effort with the CHTS. The CHTS used random-digit dialing to reach households, using coastal county telephone prefixes. Initially, the CHTS saw high response rates but was inefficient, meaning that many non-angling households were contacted for every angling household that answered. Because the CHTS did not contact non-coastal county anglers, they were captured in the on-site survey component of the survey and the ratio of coastal to non-coastal anglers was used to increase the effort obtained from the CHTS. Several trends have rendered the CHTS less efficient and potentially less reliable over time. Telephone prefixes are now portable, such that a person who first got her telephone number in Kansas may now be living and fishing in Florida. Prefixes can no longer be relied on to indicate a coastal county resident. Moreover, telephone response rates have fallen dramatically with the almost universal use of caller ID. Also, the CHTS relied on land-line telephones and the majority of US households are now wireless only. Wireless-only households have different demographic characteristics than do land-line households, and NMFS can no longer be certain that the CHTS provides unbiased or efficient estimates of effort. NMFS investigated several methods to replace the CHTS and chose a mail survey (FES) that includes a small reward and multiple mailings as is standard practice for such surveys.

The task of the MRIP Calibration Review Panel was to evaluate the performance of a new calibration model developed by F. Jay Breidt, Teng Liu, and Jean D. Opsomer of Colorado State University that permits conversion of telephone-survey effort to mail-survey effort and vice versa. NMFS has undertaken concurrent mail and telephone surveys for 2015-2017 to which the calibration model has been applied. One and one-half years of the concurrent survey evaluation has been completed at the time of this review.

Review Activites:

Review of the MRIP FES Calibration took place at the Sheraton Silver Spring, Silver Spring, MD on June 27-29, 2017.

Prior to the meeting, I reviewed documents that were provided for us on a Confluence web site two weeks before the meeting. For the first two days of the meeting, there was a series of presentations that covered issues related to the two terms of reference and five sub-terms of TOR1. On Wednesday, the reviewers requested further clarification of the presenters on several issues relating to model specification. Meetings included questions from the Panel, the audience and web participants. The Panel began work on the report Thursday. Reviewers contributed equally to the discussions. On Friday July 7, Dr. Rago conducted a conference call to further discuss TOR 2. Upon my return home, I re-read the documents, reviewed the presentations and rapporteurs' notes, and obtained several other references to help me clarify my understanding of the calibration model. These are listed in the references section of this document. I participated via email in further edits of the Panel report prior to its submission.

A very detailed review of activities is included in the Panel Review (Appendix 4).

Summary of findings for each TOR wherein weaknesses and strengths are described, with conclusions and recommendations in accordance with terms of reference:

Calibration Model Accounting for a Recreational Fishery Survey Design Change

TOR1. Evaluate the suitability of the proposed model for converting historical estimates of private boat and shore fishing effort produced by the CHTS design to estimates that best represent what would have been produced had the new FES design been used prior to 2017.

The Panel concurred that is TOR was met.

1a) Does the proposed model adequately account for differences observed in the estimates produced by the CHTS and FES designs when conducted side-by-side in 2015-2016?

I concur with the Panel's statement under TOR 1a and agree with the statements included in the Panel Review Report (Appendix 4).

It is concerning that there is a 4 to 11 fold difference in estimated trips between the CHTS and the FES and this begs an explanation.

The National Academy of Sciences (2017) and the American Statistical Association have both reviewed the FES design and agree the methodology is statistically sound. The sampling frames differ between the CHTS and the FES. The CHTS uses coastal county prefixes with random digit dialing (RDD) to contact potential angling households, while the FES uses a list of addresses of coastal state residents overlain probabilistically with the list of residences of anglers holding state licenses. The FES also gives higher selection probability to the coastal county addresses (Thereby permitting potential comparisons between the CHTS and FES strata albeit with different sampling frames). The FES is a more efficient survey because of how the angler lists are used to increase inclusion probabilities of angling households. Moreover, anglers will answer a survey differently based on the mode of contact, mail or telephone (Dillman 2014). With RDD, the angler has no prior warning that they will be asked about their fishing trips and they may also be influenced by the survey agent asking the questions. They can ask the agent for clarifications, but may not have a calendar nearby to prompt their recall on the number of trips that they took in the past two months. However, depending on when the call is received there is a chance that not all anglers in the household would be home. With the FES, the angler has time to review their calendar (if they use one) or to think about the trips that they took, and all anglers in the household have time to answer the survey. However, if the respondents have a question not included on the FAQ sheet sent with the survey, then they may mis-interpret a question. In both cases, the answers are self-reported by the angler with no external verification as to trip number or location.

Some of the differences that might occur between the surveys have been explored as predictive covariates to the model, but none were influential except, to a small degree, the increase in wireless telephone coverage over time beginning in 2000. Initially, telephone response rates were high, but with the increasing proliferation of wireless-only households and caller ID, telephone response rates have plummeted. Thus, land-line households may represent a different demographic from the target population of marine anglers that the survey seeks to contact. I am not aware if there has been a study of the demography of the anglers responding to the CHTS or the FES that might help to uncover the differences in trips reported. Please note that response bias and response rates are two different issues. Just because response rate is low does not mean that the anglers contacted differ from those not answering. A non-response survey is necessary to discover bias. However, if the CHTS is not covering the full target population and if the demographics of those who respond have different fishing characteristics, then there is cause for concern that bias might exist. Without further investigation, one is left to conjecture with no proof.

Nonetheless, the FES rests on a statistically sound sampling design with known sampling inclusion probabilities, and is far more efficient than the telephone survey at reaching an angling household. Because the response rate has been higher for mail surveys, sample size can also be larger with potential concomitant decrease in variance –thereby lessening uncertainty. Additionally, with greater sample size, the underlying distribution of number of trips per household can be better characterized.

1b) Is the proposed model robust enough to account for potential differences that would have been observed if the two designs had been conducted side-by-side in years prior to 2015 with regards to time trending biases?

I concur with the Panel's statement under TOR 1b and agree with the statements included in the Panel Review Report (Appendix 4).

Although there are studies in other fields that have tried to uncover differences between survey modes (How the survey is delivered), without actual side-by-side assessments an answer is pure conjecture. One has to assume that any trends, for example in demographic types of recreation, have been influential on participation in recreational angling and in addition, that such trends would be consistent. Although NMFS conducted a short pilot study in North Carolina for 2012-2013 on the mail survey design, there are simply no data upon which to form a conclusion. To date, none of the possible factors that are hypothesized to cause differences in effort estimates between the CHTS and the FES has been shown to account for the differences seen in trips reported.

After returning from the Panel meeting, I have been wondering if the MRIP team have any data to explore the role of "gatekeeper" in the telephone survey. The gatekeeper is the person who answers the phone. I have been wondering whether such persons answered for themselves only, which could account for the difference. I don't know whether there are data to compare trips reported based on number of anglers in a household, or even if that has been done already. However, one could also hypothesize a difference if the demographic has been changing in the CHTS to older people who don't fish as often – hence the full target population is not being reached. Again, without data, all of this is pure conjecture.

1c) How does the approach used in developing the proposed FES/CHTS calibration model compare in terms of strengths or weaknesses with other potential approaches?

I concur with the Panel's statement under TOR 1c and agree with the statements included in the Panel Review Report (Appendix 4).

The advantage to the current calibrations model is the use of a modified Fay-Herriot small-area approach which is widely respected by statisticians (Datta et al., 2005, among others). The statisticians who developed the calibration model are skilled in this approach; the model has well-defined statistical properties, and can be used to evaluate potential factors that might explain differences in the number of reported trips. The calibration team has also derived a new way of formulating the variance estimators for the model that now allows for the use of off-the-shelf software. Having readily available, tested software saves time and lowers costs of producing estimates of effort and variance for either forward or back projecting units of effort in FES or CHTS equivalents.

The Panel also discussed other types of models that could be used for calibration. Even though this was not the task assigned to us in this review, the use of other models would have value. Dr. Sullivan suggested that the team look into the use of a Bayesian approach. That had been attempted by the Calibration Team with less than good success, but may be better implemented by different software and modeling approaches. The value of other models is that they may validate the difference seen in the two surveys or may be better able to retrieve explanatory variables that

drive the differences. I would endorse this approach but think that the differences are more probably a result of problems in telephone coverage of the full target population, having better access to all household anglers through a mail survey, and a fundamental difference in how people respond to mail and telephone surveys. Hence, I don't think there is an easy answer to understanding the effort differences.

1d) Does the proposed calibration model help to explain how different factors would have contributed to changes in differences between CHTS and FES results over time?

I concur with the Panel's statement under TOR 1d and agree with the statements included in the Panel Review Report (Appendix 4).

The calibration model developed by Breidt, Teng and Opsomer permits the inclusion of covariates that can be used to uncover factors that account for differences in the effort estimates from the FES and CHTS. To date, there is no single factor that thoroughly accounts for the changes in the number of trips provided by the telephone survey. Trends in non-responses for telephone have not been explicitly modeled by factors other than the increase in wireless coverage that began in 2000. Even so, this factor accounts only for five percent of the modeled differences between the FES and CHTS projected back through time. It is important to note that only one year and one-half of three years of the side-by-side testing has been completed at this time. The model includes an "irrational" factor that the models have been unable to attribute to a known factor despite extensive efforts to uncover the reason for the different estimates.

The calibration model is detailed to the state and wave level, and even with such a short side-by-side survey has fit the data well, in part because of the small-area estimators that underlie the model. It will be important to test the stability of the model parameters as the next half of the data is included. The Panel has suggested that the model be cross validated with that new data, and I concur that will be an important test of the model. The model will not be used on the survey data until the three-year period of data collection in completed, and this will give the statisticians time to fine tune the model.

1e) Is it reasonable to conclude that revised 1981-2016 private boat and shore fishing effort estimates based on the application of the proposed FES/CHTS calibration model would be more accurate than the estimates that are currently available? Does evidence provided for this determination include an assessment of model uncertainty?

I concur with the Panel's statement under TOR 1e and agree with the statements included in the Panel Review Report (Appendix 4).

I was rather surprised by the wording of this TOR subcomponent. It seeks the panel to evaluate accuracy of the estimates, when in fact that is not possible. It led me to think that there is confusion about the type of data that are provided by offsite surveys such as the CHTS or FES. Anglers self-report their trip numbers in these surveys and there is no external validation of effort. The anglers' trips are not counted while they are fishing or when they complete their trip on site, but rather they must recall the number of trips that they took within the past two months. Many anglers do not keep a diary, perhaps some keep a calendar, but there is a possibility that these trips

are mis-remembered. While there may be little motivation to exaggerate fishing effort, a variety of factors can result in the reported trips differing from the actual number of trips taken and this type of problem is well documented in the survey literature. To determine accuracy, a validation study would need to be devised that paired an onsite validation with the offsite survey. For such a large scale survey effort, this would be difficult and very expensive.

The calibration model does provide an estimate of uncertainty even though it doesn't explain the differences in the estimates. I believe that this is the best approach at this time with the data available.

Because the effort estimate is combined with CPUE from the APAIS to estimate catch, there is an advantage to the fact that the FES is more efficient, statistically sound, and can potentially have a larger sample size. A larger sample size (more respondents) often results in smaller variance and better characterization of the effort distribution and, thus may result in less uncertainty when combined to produce estimates of catch.

TOR2. Briefly describe the panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.

I concur with the Panel's statement under TOR 2 and agree with the statements included in the Panel Review Report (Appendix 4). The Panel took this TOR very seriously, we provided a detailed response to the TOR, and I will not repeat what we presented in the report.

Having just completed the NAS MRIP Review, and having participated heavily in reviewing the FES and APAIS methodologies, I was very familiar with the issues underlying the review of the calibration model. Even so, I wished that more material had been available prior to the meeting to inform me and fellow panelists of the previous reviews and workshops that address the issue for this panel review. Moreover, the statisticians were not aware of the TORs until shortly prior to the meeting and had less time to prepare their presentations to address the TORs directly. The statisticians on this project are among the best in the world and they were able to provide us with much information in a short period of time. However, we did not see detailed information on their initial explorations into model choice that would have led to a more productive meeting. They explained that they had tried other models that weren't as good as the Fay-Herriot approach and on the second day, they provided results of an Akaike Information Criteria test of different model configurations including the simple ratio estimator with error. Because there is a serious issue that will potentially affect allocation between fishing sectors given the new estimates, it was important that we had as much information as possible. The Panelists and statisticians understood the importance of this issue and did extra work to fill in gaps that were a consequence of this. For example, I went over the ASA evaluation that I hadn't seen previously, and amended my reading with other statistical papers on the Fay-Herriot approach.

I commend the presenters, panelists, and coordinators with a very professionally run meeting. Panelists were fully engaged, and the presenters very responsive to our questions, provided responses within 24 hours. The Confluence website was easy to access and made my work

much easier than other CIE websites I have used. The conference room was well equipped and located conveniently. It was easy to see the presentations and hear the discussions. Dr. Rago did an outstanding job as Panel chairperson.

During the meeting, I brought up my concerns with communication of the calibration model and why the survey method was being changed, especially to the angling public. In my experience over 30 years with recreational angling surveys, I know that the estimates are only as good as the data and that the quality of the self-reported data especially will rest on the angler's belief in the legitimacy of the survey itself. I have found that conveying ideas such as a random sample to the lay public is challenging, even to a trained communicator. These ideas are not simple and the FES is complex. A recent article in the Virginian Pilot by our local outdoor writer complained that NMFS was transitioning to an old-fashioned survey method, and asked why didn't they just use smartphones (Tolliver, 2017)? I expect that the MRIP team will find challenges in conveying to the average angler that the mail survey is superior because of its probability basis compared with a volunteer smartphone survey that has unknown inclusion probabilities and sampling frame. I was contacted after the meeting by Gordon Colson who provided me with additional information on the MRIP communication approach. Nonetheless, the difficulty of the task of communicating to the angling public shouldn't be underestimated.

Communication to stock assessment scientists and fishery managers is also vital as they transition exclusively to the FES. The marked difference in effort estimates between the FES and CHTS has ramifications on assessments of stock status, on how to knit the time-series together, and on the allocation of catch between the commercial and recreational sectors. In some fisheries, the initial impact will be large and possibly disruptive. The MRIP communication to these two groups will also rely on the difficult task of conveying concepts that underlie survey sampling, an area of statistics not commonly taught even to quantitative scientists.

Appendix 1: Bibliography of materials provided for review

Transition Plan for the FES:

https://www.st.nmfs.noaa.gov/Assets/recreational/pdf/MRIP%20FES%20Transition%20Plan%20FINAL.pdf

Report recommending the FES to replace the CHTS: *Finalize Design of Fishing Effort Surveys* (https://www.st.nmfs.noaa.gov/pims/main/public?method=DOWNLOAD_FR_PDF&record_id=1179)

2015 Benchmarking Progress Report:

https://www.st-test.nmfs.noaa.gov/Assets/recreational/pdf/2015_FES_Progress_Report-20161115.pdf

Report on FES/CHTS Calibration Model:

BACKGROUND INFORMATION

(1) Presentations at the review

- Introduction – Paul Rago
- MRIP Fishing Effort Survey – Rob Andrews
- Importance of calibrated catch for fishery stock assessments – Richard Methot
- Importance of Calibrated Catch for Fisheries Management – Andy Strelcheck
- Calibrating survey estimates over time – Jean Opsomer
- A Calibration Methodology for CHTS to FES
- Transition – Jay Breidt
- Day One Review – Paul Rago
- Follow Up on Comments for " A Calibration Methodology for CHTS to FES" – Jay Breidt

(2) Other Papers that I Read

Datta, G.S., Rao, J.N.K., and Smith, D.D. 2005. On measuring the variability of small area estimators under a basic area level model. Biometriks 92-1: 183-196.

Dillman, D.A., Smyth, J.D. and Christian, L.M. 2014. Internet, Phone, Mail, and Mixed-Mode Surveys: a tailored design method. 4th Edition, Wiley.

Fay III, R.E. and Herriot, R.A. 1979. Journal of the American Statistical Association, Vol. 74, No. 366 (Jun., 1979), pp. 269-277.

NAS. 2017. Review of the Marine Recreational Information Program (MRIP). National Academy Press. Washington, D.C.

Tolliver, J. 2017. How many fish are really in the ocean? Some congressmen think federal fisheries can do a better job of finding out. Virginian Pilot, April 25, 2017. https://pilotonline.com/news/local/environment/how-many-fish-are-really-in-the-ocean-some-congressmen/article_dfc2f052-dab8-590c-829a-5d510dd8e983.html.

Appendix 2: A copy of this Statement of Work

Statement of Work

National Oceanic and Atmospheric Administration (NOAA)

National Marine Fisheries Service (NMFS)

Center for Independent Experts (CIE) Program

External Independent Peer Review

Calibration Model Accounting for a Recreational Fishery Survey Design Change

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards. (http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).

Further information on the CIE program may be obtained from www.ciereviews.org.

Scope

The Office of Science and Technology requests an independent peer review of a calibration model proposed for use in revising statistics produced by surveys of marine recreational fishing effort on the Atlantic coast and in the Gulf of Mexico. This calibration model is considered by the Marine Recreational Information Program (MRIP) to be very important to adjust historical time series of recreational effort and catch estimates in order to account for biases in past sampling and estimation methods that have become apparent with the development of a new, more statistically sound method. The calibration model is intended to account for past biases in private boat and shore fishing effort estimates that have resulted from the continued use of a legacy random-digit-dial telephone survey design that has degraded over time and will be replaced with the implementation of a new mail survey design (the "Fishing Effort Survey", or FES) in 2018.

Calibration Model for the Fishing Effort Survey

In 2015, MRIP formed a Transition Team to collaboratively plan a transition from a legacy telephone survey design to a new mail survey design for estimating private boat and shore fishing effort by marine recreational anglers. Since 2008, MRIP had conducted six pilot studies to determine the most accurate and efficient survey method for this purpose on the Atlantic and Gulf coasts. The most recent study, conducted in four states in 2012-2013, compared a new mail survey design with the Coastal Household Telephone Survey (CHTS) design that has been used since 1979. MRIP subjected the final report from the pilot project to external peer review in 2014 and certified the new survey design, called the Fishing Effort Survey (FES), in February 2015 as a suitable replacement for the CHTS. The FES is much less susceptible to potential sources of bias than the CHTS because it can reach more anglers, achieve higher response rates, and is less prone to possible recall errors. The pilot project results indicated that FES estimates were substantially higher than CHTS estimates for both private boat fishing and shore fishing.

MRIP recognized the FES should not be implemented immediately as a replacement for the CHTS, and a well thought out transition plan was needed to ensure that the phase-in of the FES is appropriately integrated into ongoing stock assessments and fisheries management actions in a way that minimizes disruptions to these processes, which are based on input from multiple data sources over lengthy time series. The Transition Plan developed by the Transition Team called for side-by-side benchmarking of the FES against the CHTS for three years (2015-2017) with the development and application of a calibration model to enable adjustment of past estimates that account for biases in historical effort and catch statistics after the second year. With this timeline, revised estimates can be incorporated into stock assessments during 2018 using a peer reviewed calibration model, and new Annual Catch Limits (ACLs) can then be set in 2019 for at least some stocks.

Requirements

NMFS requires three reviewers to conduct an impartial and independent peer review in accordance with the SoW, OMB Guidelines, and the Terms of Reference (ToRs) below. The CIE reviewers shall have working knowledge and recent experience in the design of sampling surveys, the evaluation of non-sampling errors (i.e., undercoverage, nonresponse, and response errors) associated with changes to survey designs over time, and the evaluation of differences between surveys using different modes of contact (e.g., mail *versus* telephone). In addition, they should have experience with complex, multi-stage sampling designs, time series analyses, regression estimators, and small domain estimation methods. Some recent knowledge and experience in current surveys of marine recreational fishing is desirable but not required.

NMFS will provide a Chair who has experience with U.S. fisheries stock assessments and their application to fisheries management. The Chair would ensure that reviewers understand the importance of maintaining a comparable time series of marine recreational fisheries catch statistics for use in stock assessments and their application to fisheries management. The Chair will not be selected by the contractor and will be responsible for facilitating the meeting,

developing and finalizing a summary report and working with the CIE reviewers to make sure that the ToRs are addressed in their independent reviews.

Tasks for Reviewers

Pre-review Background Documents

The following background materials and reports prior to the review meeting include:

Transition Plan for the FES:

https://www.st.nmfs.noaa.gov/Assets/recreational/pdf/MRIP%20FES%20Transition%20Plan%20FINAL.pdf

Report recommending the FES to replace the CHTS: *Finalize Design of Fishing Effort Surveys* (https://www.st.nmfs.noaa.gov/pims/main/public?method=DOWNLOAD_FR_PDF&record_id=1179)

2015 Benchmarking Progress Report:

https://www.st-test.nmfs.noaa.gov/Assets/recreational/pdf/2015_FES_Progress_Report-20161115.pdf

Report on FES/CHTS Calibration Model:

This report will be provided by the contractor (via electronic mail or make available at an FTP site) to the CIE reviewers.

Panel Review Meeting

Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The meeting will consist of presentations by NOAA and other scientists to facilitate the review, to provide any additional information required by the reviewers, and to answer any questions from reviewers.

Contract Deliverables - Independent CIE Peer Review Reports

The CIE reviewers shall complete an independent peer review report in accordance with the requirements specified in this SoW and OMB guidelines. Each CIE reviewer shall complete the independent peer review according to required format and content as described in **Annex 1**. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in **Annex 2**.

Other Tasks – Contribution to Summary Report

The CIE reviewers may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. The CIE reviewers are not required to reach a consensus, and should provide a brief summary of each reviewer's views on

the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Foreign National Security Clearance

When reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for reviewers who are non-US citizens. For this reason, the reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/ and http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-

registration-system.html. The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place of Performance

The place of performance shall be at the contractor's facilities, and at the NMFS Headquarters in Silver Spring, Maryland.

Period of Performance

The period of performance shall be from the time of award through July 31, 2017. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables in accordance with the following schedule. Within two weeks of award	Contractor selects and confirms reviewers
Within four weeks of award	Contractor provides the pre-review documents to the reviewers
June, 2017	each reviewer participates and conducts an independent peer review during the panel review meeting
Within two weeks of panel review meeting	Contractor receives draft reports
Within two weeks of receiving draft reports	Contractor submits final reports to the Government

Appendix 3: Panel membership or other pertinent information from the panel review

MRIP Calibration Model Peer Review Workshop

Sheraton Silver Spring Hotel

Silver Spring, MD

June 27-29, 2017

ATTENDANCE LIST

#	NAME	AFFILIATION
1	Paul Rago	MAFMC SSC
2	Dave Van Voorhees	NOAA Fisheries
3	John Foster	NOAA Fisheries
4	Ali Arab	Georgetown University
5	Rob Hicks	College of William and Mary
6	Cynthia M. Jones	Old Dominion University
7	Richard Cody	NOAA support ECS
8	Teng Liu	Colorado State University
9	Thomas Sminkey	NOAA Fisheries/ST1
10	Steve Turner	NOAA Fisheries SEFSC
11	Andy Strelcheck	NOAA Fisheries - SERO
12	Richard Methot	NOAA Fisheries - HQ
13	Karen Pianka	NOAA Fisheries – ST1
14	Lauren Dolinger Few	NMFS ST1
15	Chris Wright	NMFS - SF
16	Sabrina Lovell	NMFS ST
17	Patrick Lynch	NMFS ST
18	Melissa Karp	NMFS ST
19	Toni Kerns	ASMFC
20	Steve Ander	Gallup
21	Tommy Tran	Gallup
22	Melissa Niles	Fifth Estate/MRIP CET
23	Yong-Woo Lee	NOAA - Fisheries
24	Jay Breidt	Colorado State University
25	Jean Opsomer	Colorado State University
26	Rob Andrews	NOAA Fisheries
27	Ryan Kitts-Jensen	NOAA Fisheries
28	Fred Serchuk	SAFMC SSC
29	Jason McNamee	ASMFC
30	Patrick Sullivan	Cornell/NEFMC
31	Jason Didden	MAFMC
32	Daemian Schreiber	NMFS HQ
33	Laura Diederick	NOAA Fisheries

1

Appendix 4. Amended Panel Report to include text body only

Summary Report
Marine Recreational Information Program (MRIP)
Fishing Effort Survey (FES) Calibration Review

Calibration Model Review Meeting
June 27-29, 2017
Sheraton Hotel
Silver Spring, MD

July 14, 2017

Draft #4

Panel Members

Paul Rago (Chair)
Ali Arab
Robert Hicks
Cynthia Jones
Jason McNamee
Fredric Serchuk
Patrick J. Sullivan

2

Table of Contents

Executive Summary

A primary objective of the Marine Recreation Information Program (MRIP) is the improvement of the statistical basis of methods for estimating catches of recreationally caught fish in the coastal US. MRIP has implemented a new program for estimating fishing effort that relies on a mail-based survey rather than a historical telephone survey. This report summarizes a technical review of a calibration model to interrelate estimates of recreational fishing effort derived from the Coastal Household Telephone Survey (CHTS) with the Fishing Effort Survey (FES). The FES uses a mail survey and national angler registry. A panel of seven independent scientists met with consultant statisticians and MRIP staff to review a proposed methodology that could express historical estimates of fishing effort in terms of the new FES. A side-by-side experiment of the two methods, conducted in 2015 and 2016, served as the basis for this review.

The proposed methodology builds upon known properties of the CHTS and FES sampling designs, and an extensive time series of historical data. The calibration model relies on standard and highly-regarded methodology known as the Fay-Herriot method for small area estimation. Alternative modeling approaches might have been considered, but the proposed method was reasonable and scientifically-defensible. The authors are commended for introducing several innovations to estimate variances and to achieve analytical consistency. The final estimators have desirable properties and can be implemented with readily available software. The proposed model was considered an elegant approach for dynamic predictions of recreational fishing effort. Particularly notable was the property that allowed for forward and backward estimation by alternate survey modes (i.e., CHTS vs FES). The proposed method preserves design aspects of historical and current surveys and incorporates important differences among states, waves (i.e., two-month calendar periods) and fishing modes. The processes of model identification and variable selection (i.e., consideration of potential predictive covariates) were well done.

The Panel expressed concern on several topics, none of which was considered as sufficient to preclude implementation of the Fay-Herriot model. Comparison of estimates of effort derived from the side-by-side CHTS and FES surveys (2015 and 2016) resulted in large differences (2 to 11-fold). While many hypotheses were considered that might account for these differences, data analyses and the proposed model revealed no single hypothesis (or covariate) was sufficient. Further refinement of the modeling approach, particularly when the results of the 2017 side-by-side experiment are available, is recommended. Refinements include further simulation testing and cross-validation comparisons with the first two years of data. As more information is acquired about the FES there may be additional opportunities to consider alternative models for calibration. Given the importance of such changes for many stock assessments and management decisions, future modifications must be able to demonstrate significant advantages over the proposed small-area estimation model prior to consideration for implementation. The Panel recommended additional efforts to improve communication of these results to scientists, statisticians, fishery managers, and the general public. Each will require varying levels of detail. The Panel also suggests that renewed attention be given to the recommendations of two previous NAS reviews of the recreational statistics programs.

1. Introduction

1.1 Background

The Review Panel for the MRIP-FES Calibration Model Review met from June 27 to June 29 to review a statistical model developed by F. Jay Breidt, Teng Liu and Jean D. Opsomer, of Colorado State University. The review committee was composed of three scientists appointed by the Center for Independent Experts (CIE): Robert Hicks, The College of William and Mary; Cynthia Jones, Old Dominion University; and Ali Arab, Georgetown University. In addition, representatives from the New England (Patrick Sullivan) and South Atlantic (Fredric Serchuk) Scientific and Statistical Committees, and the Atlantic States Marine Fisheries Commission (Jason McNamee) served on the review panel. The meeting was chaired by Paul Rago as a member of the Mid-Atlantic Fishery Management Council Scientific and Statistical Committee.

The panel reviewed supporting documentation and presentations prepared by MRIP staff, led by Dave Van Voorhees, and their contractors from the Department of Statistics at Colorado State University. John Foster, Ryan Kitts-Jensen, and Richard Cody of MRIP acted as rapporteurs. Other staff from the Office of the Science and Technology, notably Karen Pianka, assisted in the handling of documents via a web-based application. Jason Didden of the Mid-Atlantic Fishery Management Council provided support for the webinar. Approximately 35 people participated in the open sessions of the meeting. The meeting followed the agenda in Appendix 2 with respect to the sequence but not necessarily the timing of the events. Adjustments were made for differences in the duration of presentations and follow-up questions.

1.2 Review of Activities

About ten days before the meeting the panel was given access to a comprehensive working paper summarizing the proposed statistical model. Prior the meeting, the chair met with the presenters and MRIP staff via a conference call to discuss the scope of the contributions, presentation format and draft agenda. All supporting documents and presentations were made available to reviewers via a web-based application known as Confluence. In addition, the MRIP staff added a web page to their site that provided members of the public and other managers with access to key papers and presentations. The meetings were broadcast via webinar with able assistance of Jason Didden of the Mid-Atlantic Fishery Management Council. Mr. Didden also managed all of the in-room computer and audio visual equipment.

The meeting opened on the morning of Tuesday June 27, 2017, with welcoming remarks and comments on the agenda by Van Voorhees and Rago. Participants and audience members introduced themselves. Following introductions, sessions on June 27 were devoted to presentation and initial discussions of five agenda topics. Robert Andrews provided an overview of the transition from the fishing effort surveys based on a Coastal Household Telephone Survey (CHTS) to the Fishing Effort Survey (FES), based on a mail survey. Richard Methot addressed the importance of properly calibrated effort for estimation of catch in stock assessments. Andy Strelcheck addressed the importance of catch information as a basis for fisheries management policies and decisions, such as allocation. Jean Opsomer provided an overview of the challenges of

applying calibration methods to historical time series. Jay Breidt led the presentation of the proposed statistical calibration model.

Each presentation was followed by a question and answer period by panel members and as appropriate, by other meeting attendees. Questions from web participants were also addressed at opportune times. A formal public comment period was reserved on each day of the meeting.

The Panel met in closed session at the end of each day to discuss the day's presentations, progress toward answering the agenda, and to make plans for the following day.

Follow-up discussions on the first day presentations were held on Wednesday June 28. The Panel requested additional data and clarification from the presenters, including greater details on the model results. Day two began with an overview of the activities of Day One and an overview of the day's work plan. Most of the Panel's efforts were devoted to questions on the statistical calibration model. Material provided by Jay Breidt and colleagues enhanced the Panel's understanding of the model and its performance. A short presentation by Paul Rago used the results of model predictions to compare results over states and fishing modes (i.e., shore vs private boat).

Day Two also included a formal public comment period and an initial summary of the Panel's findings. This was done to ensure that all participants were aware of the general outcomes of the review. The Panel stressed that this summary was not to be considered a consensus report. Instead it represented a summary of the perspectives of the Panel.

Following the initial presentation of findings, the Panel met in closed session to begin writing the Summary Report. Day Three consisted of a half day meeting for Panelists only. The purpose of the meeting was to summarize the various viewpoints herein with respect to the Terms of Reference.

The Panel completed drafting this Summary Report by correspondence, evaluating each ToR. The Chair compiled and edited the draft Panel Summary Report, which was distributed to the Panel for final review before being submitted to the MRIP. Each Panelist also provided an independent summary of their perspectives and as appropriate, with details on potential improvements to the calibration model and its application. Individual panelist reports for CIE participants were sent to the Center for Independent Experts for initial editing for completeness. Reports of Panelists supported directly by the Agency via contract were sent to the Chair. All reports were made available to MRIP staff for fact checking but were not altered for content.

The Panel agreed that scientific and statistical analyses conducted by the presenters were thorough, statistically sound, and innovative. Specific comments on the details of the analyses are provided below.

2. Review of MRIP FES Calibration Model

2.1 Synopsis of Panel Review

The Panel commented that the proposed methodology builds upon known properties of the existing sampling design, the proposed new method, and extensive time series of historical data. A review of calibration approaches in other disciplines revealed no comparable attempts to adjust a historical times series forward or backward in time in response to new information from a side-by-side comparative surveys. The proposed model was considered to be an elegant approach for dynamic predictions of recreational fishing effort. Particularly notable was the property that allowed for forward and backward estimation by alternate survey modes (i.e., CHTS vs FES). Notably, the proposed method preserves design aspects of historical and current surveys and incorporates important differences among states, waves (i.e., two-month calendar periods) and fishing modes. The Panel acknowledged the extensive exploratory data analyses on model development, alternatives, and testing performed by the MRIP scientific staff and consultants. The processes of model identification and variable selection (i.e., consideration of potential predictive covariates) were well done.

Although the Panel identified several alternative modeling approaches and other candidate covariates that might have been considered, the Panel acknowledged that the proposed method was a reasonable and scientifically defensible estimation approach.

The calibration model relies on standard, well known, and highly regarded methodology. The authors are commended for introducing several innovations to estimate variances and to achieve analytical consistency. The final estimators have desirable properties and can be implemented with readily available software.

The Panel expressed concern on several topics, none of which was considered as sufficient to preclude implementation of the model. Comparison of estimates of effort derived from the side-by-side CHTS and FES surveys (2015 and 2016) resulted in large differences (2 to 11-fold). While many hypotheses were considered that might account for these differences, data analyses and the proposed model revealed no single hypothesis (or covariate) was sufficient.

Model performance was partially assessed by sensitivity analysis of specific alternative hypotheses on the distribution of the "irregular" random effect (an effort effect not accounted for explicitly in the model). However, additional simulation work may be necessary to more thoroughly test overall model performance. As additional information becomes available by the end of the 2017 side-by-side surveys, it is recommended that a series of cross-validation exercises be conducted to compare model results based on the first two years of model results. Other permutations of cross calibration comparisons may be instructive with respect to stability of model parameter estimates and prediction error induced by various data rarefaction methods. As more information is acquired about the FES there may be additional opportunities to consider alternative models for calibration. Given the importance of such changes for many stock assessments and management decisions, future modifications must be able to demonstrate significant advantages over the proposed small-area estimation model prior to consideration for implementation.

The Panel spent considerable time discussing the communication of results. It was recognized that at least three distinct audiences must be addressed: scientists and statisticians, fishery managers, and the general public. Each will require varying levels of detail without compromising the integrity of the model or its underlying principles. A "lay person's" version of the methods would be valuable

for communicating results to multiple audiences. Model results, in combination with a similar calibration exercise for the APAIS, have significant downstream impacts for assessments and management. The Panel also suggests that renewed attention be given to the recommendations concerning communications of two previous NAS reviews of the recreational statistics programs.

Despite progress in improving communication with stakeholders, the Panel is aware of important misconceptions among the angling communities regarding the transition to the new mail-based survey mode. The new MRIP website is a considerable improvement but direct, pro-active communication and dialogue with fishing groups, perhaps with downloadable podcasts, YouTubes etc. and in-person presentations to the angling community would be valuable.

2.2 Evaluation of Terms of Reference

2.2.1 Term of Reference 1

Evaluate the suitability of the proposed model for converting historical estimates of private boat and shore fishing effort produced by the CHTS design to estimates that best represent what would have been produced had the new FES design been used prior to 2017.

- The Panel concurs that this TOR and its subcomponents listed below (1a,1b, 1c, 1d, 1e) were met.

a) Does the proposed model adequately account for differences observed in the estimates produced by the CHTS and FES designs when conducted side-by-side in 2015-2016?

- The results of the side-by-side surveys are central to the development of the proposed model. The model parameterization accounts for these changes but does not provide insight into the underlying mechanisms resulting in differences in estimated angling effort.
- The new mail survey mode has advantages relative to issues of comprehensiveness of angler coverage within households, efficiency of the estimate, a better sampling frame, a more thoughtful consideration of individual angler effort, improved demographic information, better identification of fishing location, and enhanced follow-up with respondents to reduce non-response. Collectively these features are thought to yield more reliable metrics of angling effort and serve as a basis for improved understanding in the future as the new survey continues. These advantages are relevant to 2015 and onward but do not necessarily extend back to historical estimates.

b) Is the proposed model robust enough to account for potential differences that would have been observed if the two designs had been conducted side-by-side in years prior to 2015 with regards to time trending biases?

- The Panel had difficulty formulating a response to this TOR as it required conjecture about unidentified underlying causal mechanisms contributing to observed differences and hypothetical comparisons of survey mode responses in the past.

- Insufficient information was provided to inform this decision either before or during the meeting. Potential approaches were discussed but could not be implemented in the time available.
- Although the proposed model allows for inclusion of other causal mechanisms, neither the investigators nor the Panel were able to identify covariates that vary over time and meet the criteria necessary for expansion to total angling effort estimates. Furthermore, data collection procedures during the CHTS did not collect information that in retrospect (e. g., demography, gender), might have allowed such inference.

c) How does the approach used in developing the proposed FES/CHTS calibration model compare in terms of strengths or weaknesses with other potential approaches?

- The investigators conducted an extensive analysis of within-model comparisons of reduced model parameterizations using the model selection procedure known as the Akaike Information Criterion. One sub-model included a simple ratio with random effects that had much lower explanatory power. A preliminary analysis was conducted and reviewed by the Panel that corroborated the inappropriateness of the simple ratio estimator.
- Other models exist that could be used, including Bayesian Hierarchical modeling, state-space modeling, and time-varying ratio estimation. The investigators provided the panel with a summary of their experiences with some of these alternatives but the results of these comparisons were not available to the Panel. Given the responses of the investigators, the Panel concurred with the conclusion to focus on the modified Fay-Herriot approach.

d) Does the proposed calibration model help to explain how different factors would have contributed to changes in differences between CHTS and FES results **over time**?

- As noted above the causal mechanisms resulting in differences between survey estimates remain elusive.
- Raw survey data in the CHTS could be examined more carefully but it is unknown whether such data exist over a sufficient span of years to support such analyses
    - As presently configured the model is limited in terms of what can be explored but alternatives may be useful.
    - Within the existing data, there do not appear to be covariates, other than log(Population) that would explain the major differences seen between survey modes. The wireless effect captures a minor component of the contrast. The Panel and Investigators agreed that the wireless effect may be a proxy for a wide range of factors.
    - Demographic information in the CHTS would have been instructive and is essential for proper historical analyses. However, it is uncertain that such data exist over a sufficient span of years to support such analyses.

    - Consideration of spatially differentiated data that has been collected historically at a finer scale (e.g., Census tract) may yet contain information sufficient to illuminate explanatory factors related to this TOR.
- The "Gate keeper" effect has been documented as a major influence in the CHTS but a complete understanding remains difficult to identify.

e) Is it reasonable to conclude that revised 1981-2016 private boat and shore fishing effort estimates based on the application of the proposed FES/CHTS calibration model would be more accurate than the estimates that are currently available? Does evidence provided for this determination include an assessment of model uncertainty?

- No conclusions can be reached regarding the accuracy of calibrating self-reported data from one survey mode to the other. However, the Panel noted that bias in the historical CHTS may not be as large as observed in contemporary CHTS samples due to degradation of survey coverage and other factors.
- Gatekeeper, recall bias, response rate etc. indicate that the mail survey is preferred to a phone, particularly in relation to statistical and operational efficiency. This conclusion was supported by the 2006 and 2017 NRC reports, and also in a separate review conducted by the ASA.

- Response rate per se is not a problem unless differences in fishing activity differ between respondents and non-respondents

2.2.2 Term of Reference 2

Briefly describe the panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.

The following sections highlight the Panel's concerns about the peer review meeting, including preparations before the meeting and follow-up activities. The Panel recognizes the complexity of the revisions of MRIP transition process and the need to satisfy many different audiences. The following recommendations are offered in the context of constructive criticism to improve the quality of future peer-review panels. While there is some redundancy in this section with the Panel's comments in section 2.1, the text below provides additional clarification of issues and more broadly reflects the diversity of the Panelist's opinions. The text below draws heavily from comments provided by the Panelists via correspondence after the meeting. Therefore some sections below may be reflected in part or their entirety in the Panelist's individual reports.

<u>Pre-Meeting Preparations</u>

Four background documents (Section 5 , Working Papers) were provided to Panel members two weeks prior to the meeting, and all additional documents and presentation were made available to the Panel during the meeting via a web-site (i.e., Confluence). The Panel Chair provided each of the reviewers with a proposed meeting Agenda a day prior to the start of the meeting, requesting that any comments and possible changes be provided back to him before the meeting opened. As the proposed Agenda was satisfactory to all of the Panel members, no changes to the Agenda were needed.

Panelists expressed concerns about pre-meeting preparations, noting an inadequate assembly of all the pieces needed to address the terms of reference. Greater overall coordination among presenters would have been desirable to ensure that all the relevant information was covered. Additional background documents would have been useful for the review; for example, the MRIP Handbook should have been provided before to provide more information about the telephone and mail surveys. Comprehensive previous reviews of the MRIP, such as those from the National Academy of Sciences should have been brought to the attention of the Panel, not all of whom had extensive knowledge of the history of MRIP. In this context, basic details about the surveys including similarities and differences in definitions of effort (notably, the definition of angling households), questions on the questionnaires, etc. would have helped the Panel to more effectively conduct the review.

Proceedings

The review panel proceedings went smoothly. Operationally, the meeting room had sufficient space for the Panel, presenters, and meeting attendees. The sound and projection systems worked well, as did the webinar link. Representatives from the Office of Science and Technology served as Rapporteurs and provided comprehensive summary notes to the Panel.

Discussions during the 2½ day MRIP Calibration Review illuminated various issues related to the results provided in the background documents and the PowerPoint presentations. Many of the concerns involved clarification of the information provided and/or requests for additional data and analyses. Additional data, model outputs and documents were made available to the Panel during the meeting. In all cases, these requests were satisfactorily fulfilled allowing the Panel to gain fuller insight on:

- Sampling designs, strengths, and shortcomings of the telephone (CHTS) and mail (FES) survey methods, including their relative performance and sources of error.

- Development, design, statistical properties, testing, and application of the proposed MRIP FES calibration model. This included consideration of alternative modeling approaches, cross-validation of the modeling framework to examine the stability of model parameter estimates (as well as prediction errors), the sufficiency and

explanatory power of the model's covariates, and the possible underlying mechanism(s) affecting the distribution of the "irregular" random effect, which is not explicitly accounted for within the proposed small-area estimation approach.

- Potential impacts of the calibrated recreational fishing effort estimates during 1981-2016 on future stock assessments, and on subsequent fishery management policies and practices.

- Need to effectively communicate the results of the calibration work (as well as the basis and need for continuing only the mail-based survey method in the future) to various constituency groups (i.e., the recreational and commercial fishing communities; scientists; fishery managers; the lay public) so that these groups fully understand and accept the calibration results and their subsequent use in deriving recreational catch estimates for application in stock assessments and in the fishery management process.

The Review Panel acknowledged that the proposed MRIP FES calibration model developed by Breidt *et al.* was a well-suited and statistically-appropriate approach to obtain calibrated estimates of recreational fishing effort (by state and 2-month calendar quarter for shore-based and private boat anglers) during 1982-2016.

Utility of Presentations

The presentations on the implications of revised recreational catch estimates on stock assessments and on management measures and regulatory protocols were instructive, but the Panel would have appreciated more quantitative examples. For example, implications for stock assessment models could have been drawn from the previously completed scoping exercises conducted by the Northeast and Southeast Fisheries Science Centers. Similarly, the Panel noted that detailed simulation exercises would also have been instructive.

The presentation on the Fay-Herriot model was lucid and effective, but the Panel would have appreciated more details on the model components and the model building process. Also, a summary of candidate modeling approaches —and details on the process that led to the proposed model—would have been very useful. Such details, as provided on the second day of the review, were greatly appreciated.

Greater detail would have been appreciated on the survey methodologies in the phone and mail surveys. The simulation exercise was an important start, but further simulation testing beyond those conducted would have lent greater support to the applicability of the Fay-Herriot model to the CHTS vs FES calibration. Further work on simulated data sets is suggested during the third-year comparisons (i.e., when the 2017 telephone and mail survey data are fully available).

Terms of Reference

The presenters did not address the TORs directly, which made it harder for the Panel to assess the relevance of some of the information presented with regard to the TORs. Consequently, the Panel spent a substantial portion of the question/answer periods (and discussion time) on obtaining the requisite information to address the TORs. It was evident during these interactions that the model developers had conducted additional work relevant to the TORs (such as investigation of additional modeling approaches). However, because the developers were unaware of the TORs, neither the primary report nor the presentations specifically addressed the TORs. Follow-up work accomplished by the developers during the meeting and subsequently shared with the Panel gave the Panel confidence that sufficient model scoping had been performed.

The TORs presume that converting CHTS to FES is the appropriate way to standardize the MRIP effort data. However, the statistical work available for the review primarily focused on the mathematical aspects of the calibration and not on which set of estimates reflects a truer representation of fishing effort. Lacking a sufficient statistical justification for standardizing the MRIP data to the FES estimates created problems both during the review and in addressing the TORs.

TOR1e seeks the Panel's opinion concerning the accuracy of effort estimates obtained from the CHTS and the FES. The Panel understands that any survey conducted offsite of the fishery, such as mail or telephone surveys, rely on angler self-reported data which is not subject to verification. Self-reported data is subject to a variety of biases including recall problems which can result in misremembered time and number of trips. Without an external measure of fishing from an onsite survey covering the same population in space and time, angler self-reported data cannot be verified. While the Panel comments on the calibration from CHTS to FES, there is no basis to comment on accuracy of either survey.

Documentation for Meeting

It would have been helpful for the Panel to have been provided (several weeks before the review) additional background documents (available from the MRIP Team and/or the MRIP Website) to enhance a collaborative understanding by Panel members of various aspects of the MRIP program and of recent analyses using MRIP data. For example, the *MRIP Data User Handbook,* and the October 2016 report, '*Possible Effects of Calibration Scenarios on Stock Assessments Planned for the MRIP Fishing Effort Survey Transition'* would have especially useful for Panel members to have had and read before the actual peer review occurred

Prior to the presentation and discussion of the Breidt *et al.* report at the Peer Review, this report was difficult to understand for anyone other than a highly-trained statistician.

Although a more complete understanding of this report was fostered by distribution of a PowerPoint presentation a week or so before the Review Meeting (and subsequently enhanced at the meeting by direct dialogue and interaction with the authors of the paper who clarified and responded to many issues raised by the Panel), it is recommended that in any future reviews in which a highly technical paper is seminal to the crux of such reviews that efforts be made by the paper authors to present the essence of their work in a manner that facilitates full appreciation and understanding of the import of such work by educated non-specialists. This becomes especially critical when the methods/approach provided in a paper will have significant downstream effects. This matter should be recognized in the future APAIS peer review.

Ancillary Analyses

The Panel appreciated the opportunity to investigate the details of the statistical calibration/prediction model on day 2. The model and assumptions were well thought out, but the Panel needed to better understand model inputs, parameter definitions, and nuances of the Fay-Herriot model. Similarly, the Panel appreciated the opportunity to solicit more information on model development and model selection beyond what was initially available at the meeting. Panelists received model parameter estimates upon request but did not have time at the meeting to explore them fully. Access to more detailed model outputs and the estimation code in R would have been valuable.

Also, apparently, several independent data analyses existed too, separate from the model, and it would have been good to have had a presentation and some discussion on that. Exploratory analyses of the pairwise calibration data was considered useful and should be considered for summarization when the analyses of the 2017 data are conducted.

Communication

Panelists expressed concerns about the need for improved communication at several different levels:

- to the Panel prior to the meeting,
- within the various analytical components,
- to the members of the Transition Team,
- to broader audience of stake holders.

An advantage of the current review was the inclusion of several external independent experts having expertise beyond fisheries science. This helped ensure that the methods were critically evaluated and represented state of the art, but increased the burden during pre-meeting preparations to ensure that all relevant contextual documents were available and fully comprehensible. Concerns were expressed that information essential for the review was not provided at level of detail that the Panel members expected.

The transition from the MRFSS to MRIP has required a massive restructuring of the data collection procedures while maintaining a continuous time series of reliable catch data.

Continuity of data has required coordination with governmental, academic, and industry stakeholders. Likewise, the process has involved multiple experiments and survey tests to demonstrate the value of proposed changes and development of advanced calibration approaches. This review constituted one component of this transition. Despite enormous improvements in the MRIP website and availability of raw and processed data at varying degrees of resolution, the Panel recommended greater coordination among the diverse analytical groups. The complexity of the transition requires that technical reviews are both sequential and interdependent. As such the review of any single technical issue (e.g., calibration between CHTS and FES) must rely upon and recognize the conclusions of earlier Panels. In the present review, this Panel relied on the conclusions of the ASA reviewers who noted the superiority of the FES over CHTS. Independent panels of scientists rarely accept prior reviews without questioning. Indeed, this is the nature of science. Hence it essential that each Panel in future reviews be provided with a summary of the full set of previous reviews and their relationship to the current review.

There is a strong need to effectively communicate the results of the calibration work (as well as the basis and need for continuing only the mail-based survey method in the future) to various constituency groups (i.e., the recreational and commercial fishing communities; scientists; fishery managers; the lay public) so that these groups fully understand and accept the calibration results and their subsequent use in deriving recreational catch estimates for application in stock assessments and in the fishery management process. Consideration should be given to a variety of communication approaches including but not limited to public meetings, seminars, podcasts, YouTube, and use of skilled educators.

Finally, it is recommended that an updated report/timetable/chart be prepared to illustrate current progress in meeting the tasks and timelines identified in the FES Transition Plan. This undertaking should also take note of how the recommendations tendered in all previous peer reviews of the MRIP Program (including the 2006 and 2016 NAS Reviews) have been addressed.

Improvements to Future Peer Review Processes

The Panel noted that review process left little time for an intensive review of the data, the model, and the computer code used to develop the results. Such analyses are often part of a stock assessment review (e.g., SAW/SARC https://www.nefsc.noaa.gov/saw/, or SEDAR http://sedarweb.org/). In the spirit of improving future reviews, the Panel suggests consideration of more broadly based working groups based on scientific input within and outside NOAA Fisheries. In stock assessments working groups have a strong technical focus and meet several times prior to the final assessment. Working groups would have the opportunity to examine the proposed methodologies in greater detail, included detailed reviews of the data and methods, and tests with simulated data. Exchanges of code, or reliance on standard packages in stock assessments provide both quality assurance and opportunities for improvements. Moreover, the products of working groups typically assure subsequent reviewers that the products under review are comprehensive and representative of diverse viewpoints. In particular, a working-group process would document the model

building process and allay concerns of reviewers who will always wonder why a particular alternative was not considered. Having those prior decisions as a matter of record would enhance the efficiency and quality of the review process.

The Panel recognizes that this recommendation would need to be part of the overall transition from MRFSS to MRIP. Indeed, the current Transition Team process that has regular updates on progress, conversations with stock assessment scientists and various stakeholders, and plans for upcoming tasks, already includes the essential elements of a more focused working group approach. In view of the importance of upcoming technical decisions for stock assessments, managers and harvesters, the Panel strongly urges consideration of this proposal.